8.3 Describing and Analyzing Data – Overview

Measures of Center

Mean (Average)

- Simple, arithmetic average of the data.
  - Sum all numbers and divide by the sample size (n).

$$\boxed{\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}}$$

- Same calculation for the population and sample mean (just different notation).
  - Sample mean = $\bar{x}$ (pronounced "x-bar")
  - Population mean = μ (Greek letter mu)
- Mean is NOT a resistant measure.
  - This means it is heavily affected by outliers.

Example 1

Data: 1, 5, 2, 9, ~~3~~ 30

$$\bar{x} = \frac{1+5+2+9+3}{5} = 4$$

New $\bar{x}$ (with 30) = 9.4

Median (Middle)

- The middle value in an ordered list.
- Median IS a resistant measure.
  - NOT affected by outliers.

Example 2

Case 1 – Odd n

7 Obs: 10, 5, 6, 1, 3, 9, 8

Sorted: 1, 3, 5, 6, 8, 9, 10

Med = 6

Case 2 – Even n

8 Obs: 10, 5, 6, 1, 3, 9, 8, 3

Sorted: 1, 3, 3, 5, 6, 8, 9, 10

$$\text{Med} = \frac{5+6}{2} = 5.5$$

Mode (Most common)

- The most frequently occurring value(s).
  - Unimodal data has one mode. → ex) Case 2, mode = 3 (twice)
  - Bimodal data has 2 modes.
  - Multimodal data has more than 2 modes.
  - Can be no modes (every value is distinct). → ex) Case 1, no mode
- This is the only measure of center that can be used with categorical data.

ex) Most common favorite color (can't average this)

Measures of Spread (Dispersion)

Range

- $\boxed{\text{Range = Max - Min}}$ ex) Case 2 → Range = 10 − 1 = 9 (Max − Min)
- Gives idea of the entire "range" of values, how much distance do they span in total.

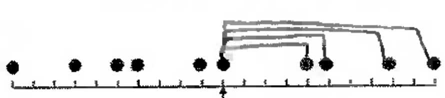
Standard Deviation

- Complex formula that measures the average distance that each data point is from the mean.

Sample Standard Deviation = $S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

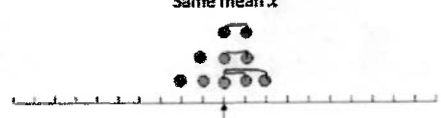
Population Standard Deviation = $\sigma = \sqrt{\frac{\Sigma(x-\mu)^2}{N}}$

(Greek letter sigma)



Larger st dev

Same mean $\bar{x}$



Smaller st dev

Using your Calculator!

Using TI-83/84 (and TI-30 XS MultiView / XIIS) to calculate mean, median, sample / population st dev.

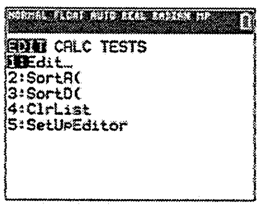
Steps for the TI-83/84	Steps for the TI-30XS MultiView	Steps for the TI-30 XIIS
1. Enter data: STAT → Edit → Enter data in L_1 *(Demo dataset: 10, 23, 4, 6, 9, 3, 15, 6)* 2. Calculate: STAT → CALC → 1-Var Stats a) List is L_1. b) Leave FreqList blank. c) Calculate!	1. Data → Enter data in L_1 2. 2^{nd} → stat → 1-Var Stats a) DATA: L1 b) FRQ: ONE c) CALC	1. 2^{nd} → STAT → 1-VAR (Enter) 2. DATA X_1 = # (scroll down) FRQ = 1 (for ALL Xs, scroll down) X_2 = # (scroll down) ... 3. STATVAR (scroll across) 4. To exit this menu: 2^{nd} → EXIT STAT → Y

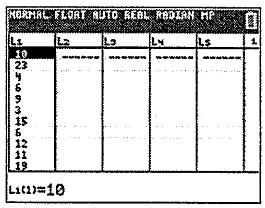
(does not give median)

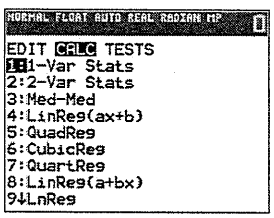
Inputs

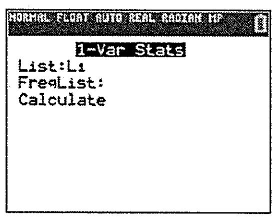
Results

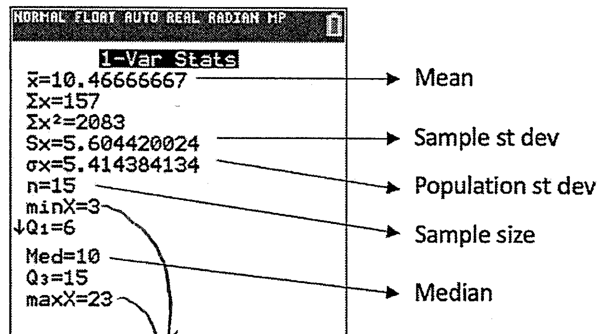
Data used here:
10, 23, 4, 6, 9,
3, 15, 6, 12, 11,
19, 10, 6, 8, 15











Example 3

Find the mean, median, mode and sample standard deviation of the following dataset.

- Data (7 obs): 35, 70, 31, 37, 65, 38, 38

Results
mean $\bar{x}$ = 44.86
sample st dev s_x = 15.72
pop st dev σ_x = 14.55
Med = 38

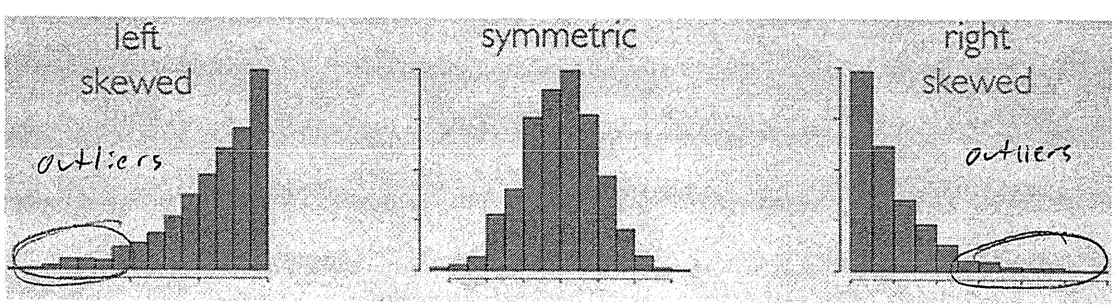
Other Considerations

Outliers

ex) 35, 37, 38, 38, 65, 70
outliers

- Data values that are extreme when compared to the rest of the data .
- Can significantly impact measures of center and spread.

Types of Distributions



Best measure of center: median | mean | median

Empirical Rule (68 – 95 – 99.7 Rule)

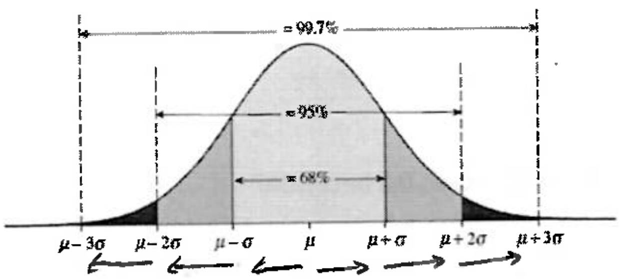
- When data is approximately bell shaped, the standard deviation allows us to make fairly accurate approximations about the locations of our data values.

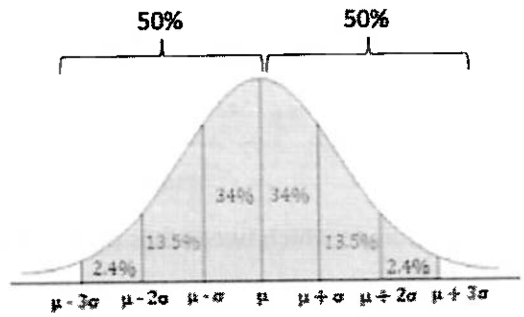
(step)

68 % of the data lies within 1 standard deviation of the mean.

95 % of the data lies within 2 standard deviations of the mean.

99.7 % of the data lies within 3 standard deviations of the mean.





- We can use these breakdowns to find probabilities within certain intervals.

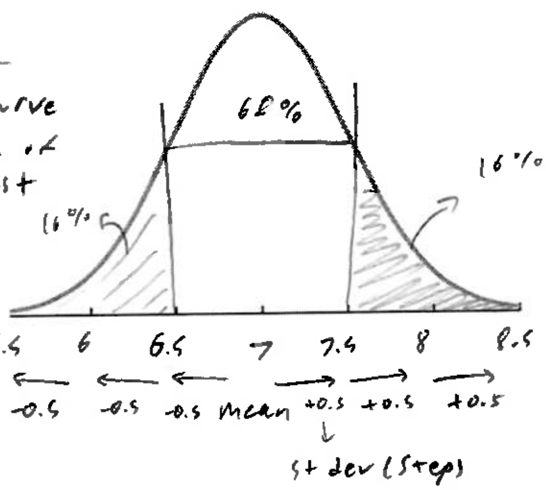
Example 4: Suppose that diameters of a new species of apple have a bell-shaped distribution with a mean of 7 cm and a standard deviation of 0.5 cm. Using the empirical rule, find the following percentages of apples with diameters that are:

Step 1

→ Draw + label curve

Step 2

→ shade curve for area of interest



a) Between 5.5 cm and 8.5 cm

Between 3 steps ⟹ $\boxed{99.7\ \%}$

b) More than 7.5 cm. → one step

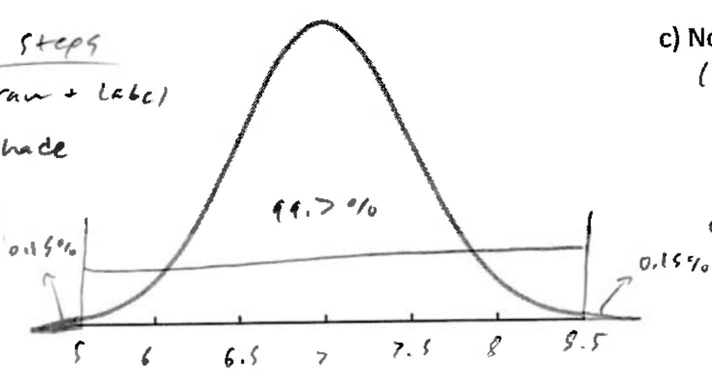
	Total		Between (Inside)	
outside →	100 %	−	68 %	= 32 %

only right → $\frac{32\ \%}{2}$ = $\boxed{16\ \%}$

Same steps

① Draw + Label

② shade



c) No more than 5.5 cm. → 3 steps

(Less)

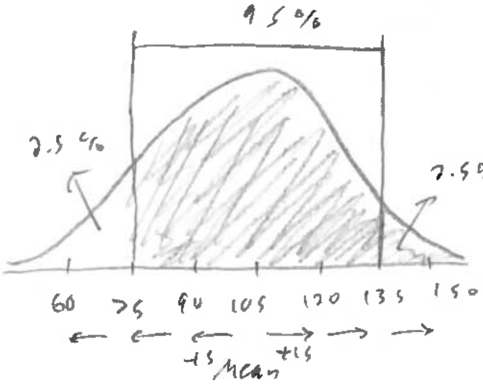
	Total		Between (Inside)	
outside =	100 %	−	99.7 %	= 0.3 %

only left side → $\frac{0.3\ \%}{2}$ = $\boxed{0.15\ \%}$

Example 5: Suppose that IQ scores have a bell-shaped distribution with a mean of 105 and a standard deviation of 15. Using the empirical rule answer the following questions:

Steps
① Draw + label
② shade

a) What percentage of IQ scores are greater than 75? → two steps



Final answer → 95% + right outside

outside → Total 100% − Between (inside) 95% = 5%

only right side = $\frac{5\%}{2}$ = 2.5%

final answer → 95% + 2.5% = 97.5%.

b) Between which two values do the middle 68% of IQ scores fall between?

Middle 68% ⇒ ± 1 step ⇒ (90, 120)

→ just wanted the two values on curve, not probabilities